ROBUST ESTIMATION METHODS FOR FIXED EFFECT PANEL DATA MODEL HAVING BLOCK-CONCENTRATED OUTLIERS

NOR MAZLINA BINTI ABU BAKAR @ HARUN

IPM 2019 15

ROBUST ESTIMATION METHODS FOR FIXED EFFECT PANEL DATA MODEL HAVING BLOCK-CONCENTRATED OUTLIERS

By

NOR MAZLINA BINTI ABU BAKAR @ HARUN

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfilment of the Requirements for the Degree of Doctor of Philosophy

January 2019

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirement for the degree of Doctor of Philosophy

ROBUST ESTIMATION METHODS FOR FIXED EFFECT PANEL DATA MODEL HAVING BLOCK-CONCENTRATED OUTLIERS

By

NOR MAZLINA ABU BAKAR

January 2019

Chair: Professor Habshah Midi, PhD
Faculty: Institute for Mathematical Research

The Ordinary Least Squares (OLS) is the commonly used method to estimate the parameters of fixed effect panel data model. However, the method is tremendously affected by the presence of outliers. In addressing the problem, we proposed new and improved robust estimators to provide resilient estimates against the most critical outlying values known as block high leverage points (HLPs). Firstly, we proposed robust panel data transformation to be performed around the MM-estimate of location as an alternative to the non-robust centering by the mean. Two robust Within Group estimators known as Robust Within Group MM (RWMM) and Robust Within Group GM (RWGM) are also proposed to be simulated under the MM-centering. Results of simulation study and real data identify RWMM and RWGM to provide more resistant and efficient estimates under MM-centering compare to the existing estimation based on median centering.

Not much research has been done on method of detecting HLPs for panel data. Hence, we have proposed Robust Diagnostic-F (RDF) to remedy the problem of masking and swamping in detecting HLPs. Simulation works and numerical examples prove that the newly proposed RDF outperforms existing methods with the lowest rates of swamping.

The existing RWGM estimator has shortcoming whereby it is based on Robust Mahalanobis Distance (RMD) based on Minimum Volume Ellipsoid (MVE) which is prone to suffer from swamping effect. To rectify this problem, the RWGM with RDF and RWGM with DRGP are developed by integrating the RDF and existing Diagnostic Robust Generalized Potential (DRGP); respectively, into the algorithm of GM-estimator. Results indicate that the performance of RWGM(RDF) estimator which uses RDF as part of its weighting scheme surpasses other methods under study.

To date no work has been focused on robust bootstrapping methods for fixed effect panel data model. Thus, bootstrapping methods known as Diagnostic Bootstrap (Boot-D) and Weighted Bootstrap with RDF (Boot RDF) are also developed to provide resistance bootstrap estimates against block HLPs. In Boot-D, a diagnostic measure is introduced to eliminate any outlier from the sampling plan whereas new re-sampling with probabilities is derived in Boot RDF. In the study, Boot RDF is found to provide robust and superior performance as confirmed by the numerical examples and simulation results.

This research also addresses the combined problem of HLPs and heteroskedastic errors for fixed effect panel data model. A two-step robust estimator called Two Step Heteroskedasticity-Outlier (TSHO) is proposed and successfully dampens both problems. This study is considered to be among the first to solve simultaneous problems of heteroskedastic and non-normal errors for panel data. Empirical evidence via simulation experiments and numerical data show TSHO to be persistent under zero or high level of contamination. Standard errors of the beta estimates are also corrected by the newly proposed heteroskedasticity- and outlier- robust standard error or HORSE estimator. Two types of robust weights are introduced in HORSE to protect against large residuals caused by block HLPs and also heteroskedasticity. In the events, simulation results indicate the lowering level of biasness by HORSE. This leads to the final conclusion that HORSE is able to produce less bias standard errors due to the robust weighting schemes introduced in its algorithm.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

KAEDAH PENGANGGARAN TEGUH BAGI DATA PANEL EFEK TETAP DENGAN TITIK TERPENCIL BERTUMPU DI BLOK

Oleh

NOR MAZLINA ABU BAKAR

Januari 2019

Pengerusi: Profesor Habshah Midi, PhD
Fakulti: Institut Penyelidikan Matematik

Kaedah Biasa Kuasa Dua Terkecil merupakan kaedah yang kerapkali digunakan untuk menganggar parameter bagi model data panel efek tetap. Walau bagaimanapun, kaedah ini sangat terjejas dengan kehadiran titik terpencil. Bagi mengatasi masalah ini, kami mencadangkan penganggar teguh baharu yang lebih baik, yang dapat memberi anggaran yang lebih teguh terhadap titik terpencil yang paling kritikal dikenali sebagai titik tuasan blok (HLP). Pertamanya, kami mencadangkan agar transformasi data dibuat menggunakan anggaran lokasi MM sebagai alternatif kepada pemusatan min yang tidak berdaya teguh. Dua kaedah Penganggar Teguh Dalam Kumpulan di kenali sebagai Penganggar Teguh Dalam Kumpulan MM (RWMM) dan Penganggar Teguh Dalam Kumpulam GM (RWGM) turut dicadangkan untuk disimulasikan di bawah pemusatan-MM. Melalui kajian simulasi dan data sebenar, RWMM dan RWGM dikenalpasti berupaya memberi anggaran yang lebih teguh serta efisien di bawah pemusatan-MM, berbanding dengan kaedah sedia ada yang menggunakan pemusatan median.

Belum ada banyak kajian yang dilakukan dalam kaedah pengesanan HLP untuk data panel. Oleh itu, kami telah mencadangkan Diagnostik Teguh-F (RDF) bagi mengatasi masalah litupan dan limpahan sewaktu mengenalpasti HLP. Kerja-kerja simulasi serta contoh berangka telah membuktikan kaedah baharu RDF ini mengatasi kaedah sedia ada dengan kadar limpahan yang paling rendah.

Penganggar RWGM yang sedia ada mempunyai kelemahan di mana kaedah ini adalah berdasarkan kepada Jarak Teguh Mahalanobis (RMD) dengan Isipadu Elipsoid Minimum (MVE) yang terdedah kepada kesan limpahan. Untuk memperbetulkan masalah ini, RWGM dengan RDF dan RWGM dengan DRGP telah dibangunkan dengan mengintegrasikan RDF dan Diagnostik Teguh Potensi Teritlak (DRGP) ke dalam algoritma penganggar GM. Hasil kajian

menunjukkan bahawa prestasi penganggar RWGM(RDF) yang menggunakan RDF sebagai sebahagian daripada skim pemberatnya melangkaui kaedah lain.

Sehingga kini, belum ada kajian yang memfokuskan kepada kaedah butstrap teguh untuk model data panel efek tetap. Oleh itu, kaedah butstrap yang dikenali sebagai Butstrap Diagnostik (Boot-D) dan Butstrap Berpemberat RDF (Boot RDF) turut dibangunkan untuk mendapatkan anggaran teguh butstrap terhadap blok HLP. Dalam Boot-D, satu ukuran diagnostik diperkenalkan untuk menghapuskan mana-mana titik terpencil daripada pelan penyampelan manakala penyampelan semula dengan kebarangkalian turut diperkenalkan dalam kaedah Boot RDF. Kedua-dua kaedah ini dapat memberikan alternatif teguh yang unggul; disahkan oleh contoh berangka dan kajian simulasi.

Kajian ini turut membincangkan masalah serentak yang melibatkan HLPs dan kesilapan heteroskedastik untuk model data panel efek tetap. Penganggar teguh dengan dua langkah utama yang dikenali sebagai HO Dua Langkah (TSHO) turut dicadangkan bagi meredakan kedua-dua permasalahan ini. Kajian ini dianggap sebagai kajian pertama yang menyelesaikan permasalahan heteroskedastisiti dan ketidaknormalan secara serentak bagi data panel. Bukti empirik melalui eksperimen simulasi dan data berangka menunjukkan TSHO sentiasa teguh di bawah pencemaran sifar atau tahap tinggi. Ralat piawai bagi anggaran beta juga turut diperbetulkan menggunakan kaedah baru yang dicadangkan iaitu penganggar ralat piawai - heteroskedastisiti titik terpencil atau ringkasnya penganggar HORSE. Dua jenis pemberat teguh diperkenalkan oleh HORSE untuk melindungi daripada ralat yang disebabkan oleh blok HLP dan juga kesan heteroskedastisiti. Dalam menangani perkara ini, hasil simulasi menunjukkan HORSE berupaya untuk mengurangkan kesan bias. Secara kesimpulannya, HORSE telah dapat mengurangkan bias ralat piawai hasil daripada skim pemberat teguh diperkenalkan dalam algoritmanya.

ACKNOWLEDGEMENTS

I would like to gratefully acknowledge all of the people whose support and encouragement have helped me finish my PhD journey.

First and foremost, I would like to express the deepest appreciation to my supervisor, Professor Dr. Habshah Midi, who has brought me a passion for research, taught me the critical thinking and scientific writing skills. I am especially thankful to her for her incredible patience in discussing my research questions, reviewing my papers and particularly this thesis. I believe that without her supervision and constant help this thesis would have not been possible.

I am grateful to the Ministry of Higher Education (MOHE) and Universiti Sultan Zainal Abidin (UniSZA) for providing me the scholarship to pursue my PhD and to the Institute of Mathematical Research (INSPEM), UPM for a great research environment.

This thesis would not have been existed without the unconditional love and endless support from my family. I dedicated this dissertation to my parents, whose love have grown me up; to my brothers and sisters for their continuous love and support; to my husband who have been always by my side.

Above all, the greatest praise be to Almighty Allah for His infinite grace, mercy and guidance.

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Doctor of Philosophy. The members of the Supervisory Committee were as follows:

Habshah Midi, PhD
Professor
Faculty of Science
Universiti Putra Malaysia
(Chairman)

Jayanthi Arasan, PhD
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Member)

ROBIAH BINTI YUNUS, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

Declaration by graduate student

I hereby confirm that:

- this thesis is my original work;
- quotations, illustrations and citations have been duly referenced;
- this thesis has not been submitted previously or concurrently for any other degree at any other institutions;
- intellectual property from the thesis and copyright of thesis are fully-owned by Universiti Putra Malaysia, as according to the Universiti Putra Malaysia (Research) Rules 2012;
- written permission must be obtained from supervisor and the office of Deputy Vice-Chancellor (Research and Innovation) before thesis is published (in the form of written, printed or in electronic form) including books, journals, modules, proceedings, popular writings, seminar papers, manuscripts, posters, reports, lecture notes, learning modules or any other materials as stated in the Universiti Putra Malaysia (Research) Rules 2012;
- there is no plagiarism or data falsification/fabrication in the thesis, and scholarly integrity is upheld as according to the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) and the Universiti Putra Malaysia (Research) Rules 2012. The thesis has undergone plagiarism detection software.

Signature: ____________________________ Date: _____________________

Name and Matric No.: Nor Mazlina Abu Bakar (GS26111)

Declaration by Members of Supervisory Committee

This is to confirm that:

- the research conducted and the writing of this thesis was under our supervision;
- supervision responsibilities as stated in the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) are adhered to.

Signature: ______________________

Name of Chairman of Supervisory Committee: Professor Dr. Habshah Midi

Signature: ______________________

Name of Member of Supervisory Committee: Assoc. Prof. Dr. Jayanthi Arasan

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

BLUE	Best Linear Unbiased Estimator
Boot RDF	Weighted Bootstrapping with Robust Diagnostic-F
Boot-D	Bootstrapping with Diagnostic Robust Generalized Potential
DRGP	Diagnostic Robust Generalized Potential
FGLS	Feasible Generalized Least Square
GM	Generalized M
GMM	Generalized Method of Moments
GP	Generalized Potential
HAC	Heteroskedasticity Auto Correlated
HLP	High Leverage Point
HOR	Heteroskedasticity Outlier Robust
HORSE	Heteroskedasticity-Outlier Robust Standard Error
HR	Heteroskedasticity Robust
LMS	Least Median Square
LTS	Least Trimmed Square
MCD	Minimum Covariance Determinant
MD	Mahalanobis Distance
MSE	Mean Square Error
MVE	Minimum Volume Ellipsoid
OLS	Ordinary Least Square
RDF	Robust Diagnostic-F
RMD	Robust Mahalanobis Distance
RMSE	Root Mean Square Error
RWGM	Robust Within Group Generalized M
RWGM(DRGP)	Robust Within Group MM with Deleted Robust Generalized Potential
RWGM(RDF)	Robust Within Group MM with Robust Diagnostic-F
RWMM	Robust Within Group MM
TSHO	Two Step Heteroskedasticity-Outliers
WBP	Weighted Bootstrap with Probabilities
WG(OLS)	Within Group with Ordinary Least Square

CHAPTER 1

INTRODUCTION

1.1 Research Background

Panel data refers to the pooling of longitudinal data, in which some units of observations, such as households, countries, firms, or nations, are followed over a number of time periods (Baltagi, 2013). Panel data analysis plays an important role in modern econometrics because its grouping structure can provide important information rather than simpler forms of data. In particular, the structure can be used to estimate models with complicated forms of heterogeneity across units or entities. Recently, there has been renewed interest in the likelihood methods for panel data, in part this is due to the emergence of new computational algorithms, such as Markov Chain Monte Carlo methods. For the past decade, there has also been an increasing trend on the use of this type of data in the research of economics and finance. The effects of globalization made it necessary to study panel data. For example, changes in an economy of a country can have significant effects to its neighbouring region due to the impact of globalization. Thus, panel data of countries in the affected regions can certainly provide more information and variability to this type of study. The uproar of Industry Revolution 4.0 makes it more essential to evaluate large panel data in order to gain critical and upfront information. Many online databases such as Osiris, Bankscope and Datastream are available online to provide the necessary and vital information. These databases have tremendously reduced the number of hours spent in collecting data. For example, panel data of firms or banks can be accessed by a few mouse clicks and the data can be downloaded almost instantaneously via the internet.

Fixed effect linear regression is one of the methods available in the econometrics. It involves the use of linear regression analysis which analyzes the relationship between a response or dependent variable and more than one explanatory variables (also known as regressor, predictor and independent variables). In general, the regression analysis helps us realize how the value of the response variable changes by changing any one of the explanatory variables in the situation that the other explanatory variables are considered to be fixed. The Ordinary Least Squares (OLS) method is often used to estimate the parameters of the fixed effect panel data model. The method minimizes the sum of squared of regression errors which are the estimate of distances between the responses predicted by the linear approximation and the observed responses in the data set. Data analyzers prefer to apply OLS due to the universal acceptance, elegant statistical properties, and computational simplicity. Unfortunately, the OLS depends on a number of fairly restrictive and often unrealistic assumptions. Among the assumptions are the normality of error distribution, independency of the explanatory variables and error terms with constant variance for all observations or homoscedasticity (Kutner et al., 2004; Baltagi, 2013; Greene, 2017). The normality assumption is often violated in the presence of one or more sufficiently outlying observations in the data set resulting in bias and unreliable estimates of the model parameters (see Montgomery et al., 2001; Gujarati, 2002; Chatterjee and Hadi, 2006; Andersen, 2008).

It is important to point out that researchers must be aware that panel data is susceptible to the occurrence of outliers. In the existence of outliers, the assumption of independent and identically distributed (i.i.d) errors for linear regression is completely violated. The least square estimate minimizes squared errors and gives high weights to outliers, causing the parameter estimates to become extremely sensitive to their presence (Maronna et al., 2006; Imon, 2017). In addressing the problem, highly advanced robust methods have been developed for linear regressions (Hampel et al., 2001; Chatterjee and Hadi, 2006; Huber, 2011). Modern robust methods are researched to find highly efficient estimators which mimic least square estimates in the absence of outliers. Typically, intensive computer simulations are required in this type of research which are now widely accessible; for example R computing by R Core Team (2013). Somehow, only limited investigations are done for regression of panel data (Wagenvoort, 1998; Bramati and Croux, 2007; Verardi, 2010; Aquaro and Cizek, 2013). The effects of outliers can be crucial for fixed effect model especially when multiple outliers occur concentrated in the time series. Any atypical observation can cause panel data to become highly contaminated due to data transformation by non-robust centering procedure (Bramati and Croux, 2007). The detection of the outliers can be very difficult due to the effects of masking and swamping (Rousseuw and Van Zomeren, 1990; Hadi, 1992; Imon, 2017). The two effects somehow can be lessened by carefully inspecting the outliers by a robust method (Habshah et al., 2009). The development of a robust outlier detection method for panel data is very important since the method can be used to determine robust weights. These robust weights can be integrated into other methods such as bootstrapping and GM-estimators to gain more efficiency and robustness in estimating robust parameters against the effects of the outliers.

Another equally important problem to be addressed in a panel linear regression model is the violation of the homoskedastic errors assumption. Heteroskedastic errors become a common problem in panel data whereby heteroskedasticity-robust (HR) standard errors become a major discussion in the econometrics literature (Arellano, 1987; Kezdi; 2004; Stock and Watson, 2006; Petersen, 2009; Imbens and Kolesar, 2016). However, no discussion is made on the simultaneous occurrence of heteroscedasticity and outlying values. The existing HR methods are only robust towards heteroscedasticity but break down in the existence of outliers (Croux et al., 2003). Thus, this thesis focuses on the effects of outliers on the fixed effect panel data model; especially *x*-outliers which lies concentrated in a few time series or known as block high leverage points (block HLPs). The coexisting problems of heteroscedasticity and block HLPs are also highlighted. Comparisons among newly developed methods and the existing ones are made. These are the paths taken by this study where robust procedures are introduced to lessen the effects of the block HLPs and also heteroscedasticity in the fixed effect panel data model.

1.2 Importance and Motivation of the Study

A small percentage of outliers is expected to present in a dataset (Hampel, 1971; Hampel, 2001). The presence further leads to a wrong conclusion of a statistical analysis. Research in this area is particularly important because many researchers are unaware of the biasness produced in the statistical analysis caused by these outliers. In panel data, outliers are frequently found to be concentrated in a few time series or also known as block concentrated outliers (Bramati and Croux, 2007). Block concentrated *x*-outliers or block HLPs can be lethal because they may introduce heavy contamination to the contaminated time series (Verardi and Wagner, 2010). The classical estimation such as the arithmetic mean is highly affected by the presence of even a single aberrant data (Maronna, 2006). Fixed effect regression requires data to be transformed around the arithmetic mean before ready to be regressed (Greene, 2017). Wrongly estimated mean will be produced when panel data are contaminated by block HLPs. As a result, the mean centering procedure will introduce more outliers into the transformed panel data. Regardless of their sources, the outliers can render least squares estimations meaningless (Kutner et al., 2004; Baltagi, 2013; Imon, 2017). Bramati and Croux (2007) have proposed the use of median centering as a robust measure to minimize the heavy contamination in the transformed data caused by the block HLPs. Median is chosen since the measure is robust and can easily be derived. However, the median centering technique is found to cause the transformed data to become nonlinear and non-equivariance (Verardi and Wagner, 2010). Moreover, median is only 64% efficient compared to mean (Maronna et al., 2006). The less efficient median will cause robust estimator to become less efficient in an uncontaminated data. These unresolved issues motivate this study to propose MM-centering by considering MM-estimate of location as the measure of central tendency. MM-estimate of location is 98% efficient and can bring back linearity and equivariance to the transformed data (Maronna et al., 2006). The MM-centering is therefore expected to provide more efficiency and thus better performance than the median centering for robust estimation in panel data.

It is highly important to identify outliers or high leverage points (HLPs) in the panel data. Once the true outliers are correctly detected, correctional measures can be taken up to remedy the problems regarding outlying values. Classically, outliers are detected by Mahalanobis Distance (MD) but the measure is non-robust and highly affected by the occurrence of HLPs (Hadi, 1992). MD simply calculates the distance of each data point from its centre mass. Theoretically, a data point with a large MD indicates outlyingness due to its large distance from the centre mass. However, MD formulation heavily relies on the non-robust arithmetic mean and covariance matrix to determine the centre mass of the data points. Thus, in the presence of outliers, both the arithmetic mean and the covariance matrix are no longer reliable, causing the centre mass to be shifted; explaining the non-robustness of MD towards outliers. From the literature, there are many different high leverage diagnostic methods such as the Robust Mahalanobis Distance (Rousseeuw and Leroy, 2003), Generalized Potentials (Imon, 2002) and Diagnostic Robust Generalized Potentials (DRGP) based on

Minimum Volume Ellipsoid (MVE) (Habshah et al., 2009). More high leverage diagnostic methods can be referred in Hoaglin and Welsch (1978), Hadi (1992) or Rousseew and Leroy (2003). However, a few evidence suggests that the existing measures which are designed to detect a single HLP may not be effective in detecting multiple HLPs (Imon, 2002; Habshah et al., 2009). Furthermore, the problems of swamping and masking become very common in the detection of the leverage values (Serfling and Wang, 2014). Swamping causes some inliers to be falsely detected as outliers whereas masking effect causes some outliers to be detected as inliers. Most robust outlier detection techniques such as robust MD and DRGP greatly suffer from swamping. The weaknesses of the existing robust outlier detection measures have inspired us to develop a new outlier detection method for panel data. Moreover, to the best of our knowledge, no research has been done in detecting outliers for fixed effect panel data model. A novel robust outlier detection method that we called Robust Diagnostic-F or RDF is formulated based on the work of Djauhari (2010). RDF is motivated by the success combination of diagnostic-robust procedure of DRGP by Habshah et al. (2009). The newly proposed robust RDF method is anticipated to be more effective in diagnosing HLPs with low swamping rate.

Several works on robust estimation have been proposed in the literature for non-panel data. One of the well-performed robust estimators which has high breakdown point and highly resistant to the HLPs is the Generalized M-estimators or GM-estimators introduced by Schweppe (given in Hill, 1977). One can refer to Simpson (1995), Mallows (1975), Krasker and Welsch (1982), and Simpson et al. (1992) for other GM-estimators. The GM-estimators are robust methods with the main aim of down weighting HLPs with large residuals. The algorithm of the most applicable GM-estimators highly depends on the HLP diagnostic methods; which fail to detect the HLPs when they occur in multiple number (Bagheri and Habshah, 2009). The issues of swamping and masking effects must be dealt with, otherwise poor results will be resulted for contaminated data (Bagheri et al., 2012). In addition, the procedures of GM-estimators are highly dependent on less efficient initial estimators such as OLS or Least Trimmed Squares (LTS) estimators. A relatively few scholars studied robustness with respect to outliers in panel data. Among them are Wagenvoort and Waldmann (2002) who proposed one-step robust estimation and Bramati and Croux (2007) who proposed Within Group GM-estimator or RWGM for the fixed effect panel data. Both methods are found to provide resistant estimates to the panel data regression. However, new methods must be proposed to derive better results, efficiencies and performance. For example, robust weights in RWGM by Bramati and Croux (2007) is determined by RMD which is known to suffer from the effect of swamping. Some inliers may be given a "0" weight and are not considered in the weighted estimation due to the swamping effect. The shortcoming of the existing RWGM-estimators has encouraged us to develop new RWGM-estimators which are more efficient and more resistant towards HLPs. In a recent development, high leverage diagnostic method, DRGP which is developed by Habshah et al. (2009) has successfully been incorporated in some robust methods such as LTS, MM and also GM-estimator by Bagheri et al. (2009). This inspire us to introduce more robust weighting schemes for RWGM by considering superior outlier detection techniques. By following similar steps of Bagheri et al. (2009), the DRGP and also the newly developed robust outlier detection method, Robust Diagnostic-F (RDF) is incorporated into RWGM to provide new robust weights for the RWGM-estimators. The methods known as RWGM with DRGP and RWGM with RDF are to provide more efficient and robust parameter estimates for the fixed effect panel data model. These methods contain precise robust weights which are determined by DRGP and RDF whilst dampening the effects of HLPs on the robust fixed effect linear regression.

Bootstrapping method is known as a powerful and popular method in parameter estimation (Hounkannounon, 2010). In bootstrapping, a parameter is measured from an empirical distribution function of the observed data; constructed by resampling with replacement (Efron and Tibshirani, 1986). However, the conventional bootstrapping techniques such as fixed bootstrapping heavily suffer from the presence of outliers (Amado et al., 2014; Imon and Ali, 2005). In the classical fixed bootstrapping or popularly known as residuals bootstrapping, ordinary least square (OLS) regression residuals are considered in its resampling plan (Salibian-Barrera and Zamar, 2002). However, when data are contaminated by HLPs, large residuals are produced by the OLS and the bootstrap samples are further contaminated by the sampling procedure (Norazan et al., 2009). Since each data have equal chance of being included in the replications of the sub-samples, more outliers may be introduced in the sub-samples of a contaminated data and causes bootstrap distribution to break down (Imon and Ali, 2002; Norazan et al., 2009). In panel data, a vast number of bootstrapping procedures found in the literature are developed to protect against heteroskedasticity (Wu, 1986; Mammen, 1993; Liu, 1988; Cameron, et al., 2008), serial correlation or/and cross sectional dependence (Goncalves, 2011; Kunsch, 1989; Liu and Singh, 1992; Kapetanious, 2008). The literature search also found that no study is available on robust bootstrapping against outliers for panel data, even though the impact of outliers in panel data is severe. However, robust bootstrapping alternatives are extensively discussed in the literature for non-panel data (Athreya, 1987; Shao, 1990; Imon and Ali, 2005; Amado et al., 2014). In robustifying the bootstrapping technique for non-panel data, Singh (1998) suggested that contaminated data are to be trimmed off but arguments arise on the percentage level of trimming. Imon and Ali (2005) then proposed Diagnostic Before Bootstrap whereby a diagnostic procedure is to be performed before bootstrapping. In this way, any outlying values are removed from the bootstraps sampling. However, the method heavily relies on the ability of the diagnostic measure to detect outliers accurately. Some good data points may be declared as outliers due to the effects of swamping (Norazan et al., 2009). It has also been suggested that harmful outliers can be excluded in the bootstrap by sampling with probabilities (Amado and Pires, 2004). In this way, the effects of swamping and masking in determining true outliers for bootstrapping can be eliminated. Inliers will receive high probabilities and have high chances to be included in the bootstrap sub-samples. On the other hand, outlying values will receive lower probabilities and hence lower chance of being included in the bootstrap. The concept is successfully implemented in Weighted Bootstrap with Probability (WBP) by Norazan et al. (2009). In WBP, each data point is allocated a probability based on the outlyingness of the data. Robust weights for the data points are then determined to provide better bootstrap estimates for contaminated data. The success of WBP in providing the best robust bootstraps estimates for non-panel data has inspired us to develop robust bootstrap estimates for panel data. The proposed method will become among the first study to propose robust bootstrapping panel data with respect to outliers, specifically the block HLPs. Thus, Diagnostic Bootstrap (Boot-D) and Weighted Bootstrap with RDF (Boot RDF) are proposed by considering the use of DRGP and RDF; respectively. The success of both the DRGP and RDF in detecting block HLPs motivates this study in determining new robust weights and hence, probabilities for the new robust bootstrapping methods. Both DRGP and RDF swamp less good points and this property is believed to increase the efficiencies of the newly proposed bootstrapping procedures.

In panel data, robust estimation is mainly discussed with respect to heteroscedasticity and/or serial correlation (for example Stock and Watson (2008) and Petersen (2009)). This is not unusual since heteroskedasticity largely occurs in panel data and in its presence, Feasible Generalized Least Squares (FGLS) is often performed to gain better efficiency than the OLS (Miller and Richard, 2018). However, when HLPs are present in heteroskedastic data, two important least square regression assumptions are now simultaneously violated. The problems require immediate solutions and some robust measures need to be identified to withstand the bad influence of both non-normal and heteroskedastic errors. The existing FGLS is based on least square and as mentioned earlier, least square estimation is highly sensitive towards HLPs (Baltagi, 2013; Stock and Watson, 2008). Robust estimators for panel data such as the existing RWGM-estimator (Bramati and Croux, 2007), followed by RWMM, RWGM(DRGP) and RWGM(RDF) have been introduced in this study for panel data. However, these methods are designed to provide resistance towards block HLPs and have never been tested against heteroscedasticity. From the literature, robust methods of panel data are proposed to withstand conditions of heteroscedasticity or outlying values separately. To the best of our knowledge, no study has been conducted in robust estimation towards both effects of heteroscedasticity and outliers. Thus, this motivates us to conduct a study; considered to be the first, in solving simultaneous problems of heteroskedastic and non-normal errors for panel data. In this manner, Two Step Heteroscedasticity-Outlier (TSHO) robust estimator is proposed and the method consists of two important steps. The first step is taken to dampen heteroskedastic errors and the second step eliminates the effects of outliers especially of block HLPs, to produce more reliable fixed effect estimates.

In the presence of heteroscedasticity, the standard error of the fixed effect parameter estimates can be biased and inconsistent (White, 1980; Mackinnon and White, 1985; Arellano, 1987). Thus, robust standard errors such as Huber/White estimators or sandwich estimators of variance by White (1980) are proposed and referred as heteroscedasticity-robust (HR) standard errors. The HR standard errors are found to be consistent and robust against heteroscedasticity but produce biasness with increasing sample size or contaminated data (Cribari-Neto and Lima, 2014). However, they are non-robust towards outliers since the HR standard errors are derived from the non-robust regression residuals. Other HR standard errors are proposed by using bootstrapping methods (Cribari-Neto, 2004; Godfrey and Orne, 2004; Wilcox, 2005; Godfrey, 2006) but again, the methods are known to be highly influenced by leverage values. It must be noted that the presence of HLPs is more crucial than the degree of heteroscedasticity; as indicated by Cribari-Neto and Zarkos (2001). A vast gap is found in the literature in providing an alternative standard error which is robust towards both heteroskedasticity and also outliers. This motivates us to propose heteroskedasticity- and outlier-robust standard error (HORSE) estimator for the fixed effect panel data. Two sets of robust weights are derived in the HORSE estimator to dampen both effects of heteroskedasticity and block leverage values.

Extensive studies have been conducted during the last decade to develop more robust estimators against outliers and/or heteroscedasticity especially for non-panel data. The tremendous development has been assisted by the abundance of high speed and cheap computing. More robust estimators must be introduced or proposed for the panel data study. In order to disseminate knowledge on robust estimation for panel data, R-computing codes for all the proposed methods are developed and published in the thesis. The codes can certainly assist other researchers in the same field to produce more advance methods in the future.

1.3 Research Objectives

The main purpose of this thesis is to develop robust estimation for the fixed effect panel data model. The robust estimators must be resistant to crucial outliers, especially those which lie concentrated in a few time series; in the *x*-direction or known as block HLPs. Some newly developed robust estimators are also expected to withstand the effects of non-normal and heteroskedastic errors. In the search of the robust estimators, other robust procedures are also developed in order to assist to the achievement of highly robust and highly efficient estimators. The foremost objectives of our research can be outlined systematically as follows:

1. To propose robust panel data transformation as an alternative to the non-robust centering by the mean.

2. To develop robust outlier detection method as a remedy to the problem of masking and swamping in detecting HLPs for panel data.

3. To establish Robust Within Group estimators which are resistant towards the presence of block HLPs in panel data by integrating successful outlier detection methods into the algorithm of GM-estimator.
4. To propose robust bootstrapping techniques for panel data and provide resistance bootstrap estimates against block HLPs.

5. To develop robust estimator and robust standard error under the conditions of non-normal and heteroskedastic errors for fixed effect panel data model.

1.4 Scope and Limitations of the Thesis

The general extent of the thesis is to proposed robust estimation for the fixed effect panel data model under the influence of outliers. In particular, the study emphases on the multiple outliers which occur in the *x*-direction and lie concentrated in a few time series. This type of outliers is known as block high leverage points (HLPs) which found to have significant effects on fixed effect parameter estimation in their presence. The effects of heteroscedasticity are also highlighted. The study would be done through the proposed robust methods based on references and coded algorithms by R-programming. Simulation experiments and applications on real data would be carry out to know the proposed methods' performances. However, this study is limited to balanced panel data with fixed effect in the presence of HLPs. This study will not cover other types of outliers nor other type of regression model. The algorithms of the proposed methods can only be proposed based on limited references of robust estimation for panel data. Monte Carlo simulations of the proposed methods will be conducted using limited sample sizes and will be based on ideal conditions and assumptions. At the same time, the simulation experiments often require a significant amount of computer time and can be expensive.

1.5 Overview of the Thesis

In accordance with the objectives and the scope of the study, the contents of this thesis are organized into nine chapters. The thesis chapters are structured so that the research objectives are apparent and are conducted in the sequence outlined.

Chapter Two:

This chapter deals with a brief literature review of the OLS estimations for the fixed effect panel data regression parameters and violations from its assumptions. The chapter reviews on the types of outliers in panel data and the existing techniques to detect them, and to understand the effects of abnormal data on the techniques. Different robust estimators, bootstrapping techniques and heteroskedasticity-robust estimators are also reviewed by highlighting the strengths and weaknesses of existing methods.

Chapter Three:

This chapter discusses the development of the robust centering procedures for the contaminated fixed effect panel data model. MM-Centering is proposed to resolve the issues by considering robust MM-estimate of location as the measure of central tendency.

Chapter Four:

A novel robust outlier detection method that we called Robust Diagnostic-F or RDF is proposed in this chapter. The newly proposed method is a combination of robust Mahalanobis Distance and diagnostic F (Djauhari, 2010) and is highly reliable in detecting the true HLPs whilst reducing the effect of swamping.

Chapter Five:

This chapter deals with the development of two robust methods which are resistant to the presence of high leverage points. The proposed robust estimators are the improvised Robust Within Group GM-estimator (RWGM) based on the DRGP and also based on the newly developed robust outlier detection method, RDF. They are known as RWGM(DRGP) and RWGM(RDF); respectively.

Chapter Six:

In this chapter, two robust bootstrapping techniques are proposed, namely Boot-Diagnostic and Weighted Bootstrap with RDF (Boot RDF). The Boot-Diagnostic is motivated by Imon and Ali (2005) and Boot RDF is proposed by incorporating new robust weights determined by the Robust Diagnostic-F (RDF) from Chapter 4.

Chapter Seven:

This chapter addresses the robust solutions towards non-normal and heteroskedastic errors. A heteroskedastic- and outlier-robust estimator is proposed in this chapter to correct the problems. The proposed method is called Two Step Heteroscedasticity and Outlier- robust estimator or TSHO which consists of two important steps (or stages) to guard against the bad effects of block HLPs and also heteroskedasticity.

Chapter Eight:

This chapter considers the computation of robust standard errors for robust estimators. Heteroskedasticty- and outlier-robust standard error or HORSE estimator is proposed by relaxing the assumptions of heteroskedasticity and provide robust weights to reduce the effects of HLPs.

Chapter Nine:

This chapter provides summary and detailed discussions of the thesis conclusions. Areas for future research are also recommended.

1.6 Flowchart of the Study

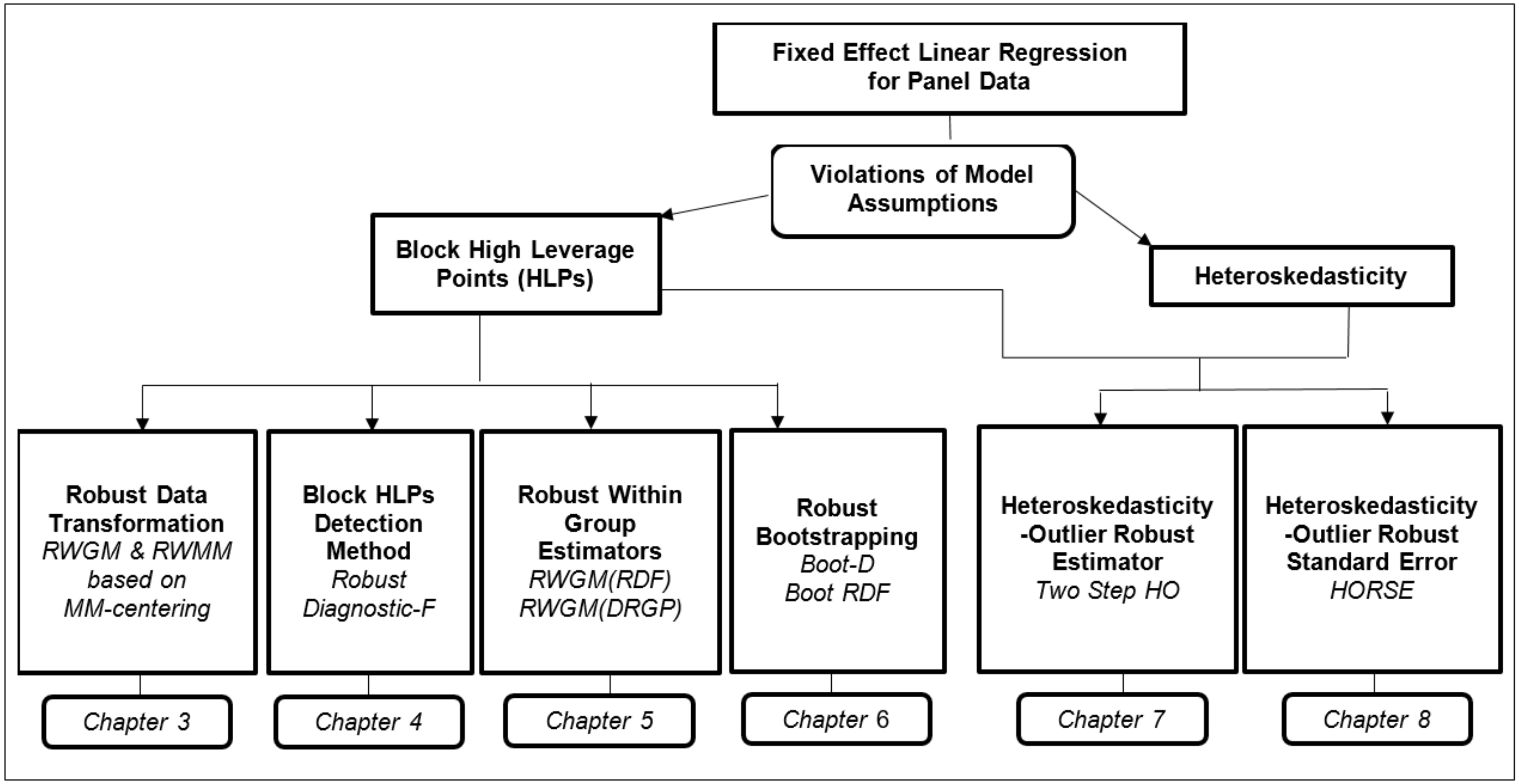

Figure 1.1: Flowchart of the study

REFERENCES

Aelst, S. V. and Rousseeuw, P. (2009). Minimum volume ellipsoid. *Advanced Review*. Volume 1: July/August 2009.

Amado, C. and Pires, A.M. (2004). Robust bootstrap with non-random weights based on the influence function, *Communications in Statistics — Simulation and Computation*. 33, 377–396.

Amado, C., Bianco, A. M., Boente, G., and Piers, A. M. (2014). Robust bootstrap: An alternative to bootstrapping robust estimators. *REVSTAT - Statistical Journal*. 12(2), 169– 197.

Andersen, R. (2008). *Modern methods for robust regression*. The United States of America: Sara Miller McCune. SAGE publications.

Andersson, M. K. and Karlsson, S. (2001). Bootstrapping Error Component Models. *Computational Statistics*. 16(2), 221-231.

Andrews, D. F. (1974). A robust method for multiple linear regression. *Technometrics*. 16, 523-531.

Aquaro, M. and Cizek, P. (2013). One-step robust estimation of fixed-effects panel data models, *Computational Statistics and Data Analysis*. 57(1), 536–548.

Arellano, M. (1987). Computing robust standard errors for within-groups estimators, *Bulletin of Economics and Statistics*. 49(4), 431-434.

Arellano, M. and Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations, *Review of Economic Studies*. 58, 277-297.

Armstrong, R. D. and Kung, M. T. (1978). Least absolute values estimates for a simple linear regression problem. *Applied Statistics*. 27, 363-366.

Arrelano, M. (2003). *Panel Data Econometrics*, Oxford: Oxford University Press.

Athreya, K. B. (1987). Bootstrap of the mean in the infinite variance case, *Annals of Statistics*. 15(2), 724-731.

Bagheri, A. (2009). Two-step robust diagnostic method for identification of multiple high leverage points, *Journal of Mathematics and Statistics*. 5, 97-106.

Bagheri, A. and Habshah, M. (2009). Robust estimations as a remedy for multicollinearity caused by multiple high leverage points, *Journal of Mathematics and Statistics*. 5(4), 311-321.

Bagheri A., Habshah, M., and Imon, A.H.M.R. (2012). A novel collinearity-influential observation diagnostic measure based on a group deletion approach, *Communications in Statistics - Simulation and Computation*. 41(8), 1379-1396.

Baltagi, B.H. (2013). *The Econometrics of Panel Data*. John Wiley and Sons, New York. ISBN: 978-1-118-67232-7.

Beaton, A.E. and Tukey, J.W. (1974). The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. *Technometrics*.16, 147-185.

Bellman L., Breitung J. and Wagner J. (1989). Bias correction and bootstrapping of error component models for panel data: theory and applications, *Empirical Economics*. 14, 329-342.

Bickel, P.J. (1975). One-step Huber estimates in the linear model. *Journal of the American Statistical Association*. 70, 428-434.

Bickel, P.J. (1965). On some robust estimates of location. *The Annals of Mathematical Statistics*. 36 (3), 847-858.

Baltagi, B. (2013) Econometric Analysis of Panel Data. 5th Edition, Wiley, Chichester.

Bramati, M. C. and Croux, C. (2007). Robust estimators for the fixed effects panel data model, *Econometrics Journal*. 10(3), 521–540.

Brownlee, K.A. (1984). *Statistical Theory and Methodology in Science and Engineering*. 2nd Edition, Krieger Pub Co., USA., ISBN: 10: 0898747481.

Butler, R.W., Davies, P.L. and Jhun, M. (1993). Asymptotics for the minimum covariance determinant estimator. *Annals of Statistics*. 21:1385–1400.

Cameron, A.C., and Miller, D.L. (2010). Robust inference with clustered data, *Technical Report*, UC Davis Department of Economics, February

Cameron, A.C., Gelbach, J.B. and Miller, D.L. (2008). Bootstrap-based improvements for inference with clustered errors, *Review of Economics and Statistics*. 90(3), 414-427.

Carroll, R.J. and Ruppert, D. (1982). Robust estimation in heteroscedastic linear models, *Annals of Statistics*, 10(2), 429-4414.

Carroll, R.J. and Welsch, A.H. (1988). A note on asymmetry and robustness in linear regression, *Journal of the American Statistical Association*. 4, 285-287.

Carvajal, A. (2000). Bootstrap confidence intervals for random effects panel data models, Working Paper Brown University, Department of Economics.

Chatterjee, S. and Hadi, A.S. (1986). Influential observations, high leverage points, and outliers in linear regression, *Statistical Science*. 1(3), 379-416.

Chatterjee, S. and Hadi, A.S. (2006). *Regression Analysis by Example*. 4th Edition. New York: Wiley.

Chou, P. and Zhou, G. (2006). Using bootstrap to test portfolio efficiency, *Annals of Economics and Finance*. 7, 217-249.

Cizek, P. (2009). Generalized method of trimmed moments, *SSRN eLibrary*.

Coakley, C.W. and Hettmansperger, T.P. (1993). A bounded-influence, high-breakdown, efficient regression estimator. *Journal of the American Statistical Association*. 88, 872–880.

Cribari–Neto, F. (2004). Asymptotic inference under heteroskedasticity of unknown form. *Computational Statistics and Data Analysis.* 45:215–233.

Cribari-Neto, F. and Silva, W. (2011). A new heteroskedasticity-consistent covariance matrix estimator for the linear regression model. *AStA Advances in Statistical Analysis.* 95:2, 129-146

Cribari–Neto, F. and Zarkos, S. G. (1999). Bootstrap methods for heteroskedastic regression models: evidence on estimation and testing. *Econometric Review*. 18, 211–228.

Cribari–Neto, F. and Zarkos, S. G. (2001). Heteroskedasticity-consistent covariance matrix estimation: White's estimator and the bootstrap. *Journal of Statistical Computation and Simulation*. 68,391–411.

Cribari–Neto, F. and Zarkos, S. G. (2004). Leverage-adjusted heteroskedastic bootstrap methods. *Journal of Statistical Computation and Simulation*. 74, 215–232.

Croux, C., Dhaene, G. and Hoorelbeke, D. (2003). Robust standard errors for robust estimators. Research report, Dept. of Applied Economics, K.U. Leuven.

Daszykowski, M., Kaczmarek, K., Vander, Y.H. and Walczak, B. (2007). Robust statistics in data analysis - a review basic concepts. *Chemometrics Intelligence Labo*ratory. 85: 203-219.

Davies, P.L. (1987). Asymptotic behavior of S-estimates of multivariate location parameters and dispersion matrices. *Annals of Statistics.* 15:1269–1292.

Davison, A.C. and Hinkley, D.V. (1997). *Bootstrap Methods and Their Application*. Cambridge University Press.

Davidson, R. and MacKinnon, J. G. (1993). Estimation and Inference in Econometrics. New York: Oxford University Press.

Davidson, and R., MacKinnon, J. G. (2004). Econometric Theory and Methods. New York: Oxford University Press.

Djauhari, M. (2010). A multivariate process variability monitoring based on individual observations. *Modern Applied Science.* 4(10).

Donoho, D.L. and Huber, P.J. (1983). The notion of breakdown point. *A Festschrift for Erich L. Lehmann*. 157 – 184.

Draper, N.R. and Smith, H. (1998). *Applied Regression Analysis*. New York:Wiley.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics*.7:1-26.

Efron, B.; Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*. 1: 54-75.

Everaert, G., and Pozzi, L. (2007). Bootstrap-based bias correction for dynamic panels. *Journal of Economic Dynamics and Control*. 31(4): 1160–1184.

Focarelli, D. (2005). Bootstrap bias-correction procedure in estimating long-run relationships from dynamic panels, with an application to money demand in the euro area. *Economic Modelling*. 22: 305-325.

Fung, W. (1993). Unmasking HLPs and leverage points: a confirmation. *Journal of the American Statistical Association*. 88: 515-519.

Gonçalves, S. (2011). The moving blocks bootstrap for panel linear regression models with individual fixed effects. *Econometric Theory*. 27:1048-1082.

Gonçalves, S. and Kaffo, M (2015). Bootstrap inference for linear dynamic panel data models with individual fixed effects. *Journal of Econometrics.* 186:407–426, 2015.

Greene, W. H. (1997). *Frontier Production Functions* in Pesaran, M. and Schmidt, P., *Handook of Applied Econometrics: Volume 2: Microeconomics*. London: Blackwell Publishers.

Greene, W. H. (2017). *Econometric Analysis*. 6th edition. Upper Saddle River. New Jersey: Prentice Hall.

Gujarati, D.N. (2002). *Basic Econometrics*. 4th edition. New York: MacGraw-Hill.

Habshah, M., Norazan, M.R. and Imon, A.H.M.R. (2009). The performance of Diagnostic-Robust Generalized Potentials for the identification of multiple high leverage points in linear regression. *Journal of Applied Statistics.* 36(5): 507-520.

Midi, H., and Bagheri, A. (2013). Robust multicollinearity diagnostic measures based on minimum covariance determination approach. *Economic Computation & Economic Cybernetics Studies & Research*, 47(4), 1-15.

Habshah, M., Bagheri, A., and Imon, A.H.M.R. (2010). The application of robust multicollinearity diagnostic method based on robust coefficient determination to a non-collinear data. *Journal of Applied Sciences*. 10(18): 611-619.

Habshah, M. (2000). Bootstrap methods in a class of nonlinear regression models. *Pertanika Journal of Science and Technology*. 8(2): 175-189.

Hadi, A. S. (1992). A new measure of overall potential influence in linear regression. *Computational and Statistical Data Analysis*. 14:1-27.

Hadi, A.S. (1992). Identifying multiple HLPs in multivariate data. *Journal of the Royal Statistical Society B*. 54: 761-771.

Hampel, F.R. (1971). A general qualitative definition of robustness. *The Annals of Mathematical Statistics.* 42(6): 1887–1896.

Hampel, F.R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association*. 69: 383-393.

Hampel, F.R. (2001). Robust statistics: a brief introduction and overview. *Research Report 24*, Seminar for Statistics.

Hampel, F.R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A. (1986). *Robust Statistics: The Approach Based on Influence Functions*. New York: Wiley

Handshin, E., Schweppe, F.C., Kohlas, J. and Fiechter, A. (1975). Bad data analysis for power system state estimation. *EEE Transactions of Power Apparatus and Systems.* PAS-94: 329-337.

He, X. and Portnoy, S. (1992). Reweighted LS estimators converge at the same rate as the initial estimator. *The Annals of Statistics*. 20(4): 2161–2167.

Herwartz H. (2006). Testing for random effects in panel data under cross-sectional error correlation: a bootstrap approach to the Breusch Pagan Test. *Computational Statistics and Data Analysis*. 50: 3567-3591.

Herwartz H. (2007). Testing for random effects in panel models with spatially correlated disturbances. *Statistica Neerlandica*. 61(4): 466-487.

Herwindiati, D.E., Djauhari, M.A., and Mashuri, M. (2007). Robust multivariate outlier labeling. *Communications in Statistics - Simulation and Computation*. 36(6): 1287–1294.

Hill, R.W. (1977). *Robust Regression When There Are Outliers in the Carriers*. Unpublished Ph.D. thesis. Harvard University, Boston, MA.

Hoaglin, D. C and Welsch, R. E. (1978). The hat matrix in regression and ANOVA. *American Statistician.* 32: 17-22.

Hodge, V.J. and Austin, J. (2004). A survey of outlier detection methodologies. *Artificial Intelligence Review*. 22 (2): 85-126.

Hounkannounon, B. (2010). Bootstrap for panel regression models with random effects. Unpublished Manuscript. Department of Economics, Universitè de Montrèal.

Huber, P.J. (1964). Robust estimation of location parameters. *Annals of Mathematical Statistics.* 35: 73–101.

Huber, P.J. (2004). *Robust Statistics*. Wiley: New York.

Huber, P. J. (1973). Robust regression: asymptotic, conjectures and Monte Carlo. *The Annals of Statistics*. 1: 799-821.

Imbens, G.W., and Kolesar, M. (2016). Robust standard errors in small samples: some practical advice. *Review of Economics and Statistics*. 98: 701-712

Imon, A.H.M.R. (2002). Identifying multiple high leverage points in linear regression. *Journal of Statistical Studies*. *Special Volume in Honour of Professor Mir Masoom Ali*. 3: 207–218.

Imon, A.H.M.R. (2005). Identifying multiple influential observations in linear regression. *Journal of Applied Statistics*. 32: 929-946.

Imon, A.H.M.R. (2009). Deletion residuals in the detection of heterogeneity of variances in linear regression. *Journal of Applied Statistics*. 36(3): 347-358.

Imon, A.H.M.R, and Ali, M.M. (2005). Bootstrapping regression residuals. *Journal of Korean Data and Information Science Society*. 16(3): 665-682.

Imon, A.H.M.R (2017); Introduction to Environmental Statistics Nandita Prakash Dhaka.

Kamruzzaman, M.D. and Imon, A.H.M.R. (2002). High leverage point: another source of multicollinearity. *Pakistan Journal of Statistics*. 18:435-448.

Kapetanios, G. (2004). A bootstrap procedure for panel datasets with many cross-sectional units. Working Paper No. 523, Department of Economics, Queen Mary, London. Forthcoming in Econometrics Journal.

Kapetanios, G. (2008). A bootstrap procedure for panel datasets with many cross- sectional units. *Econometrics Journal*, 11(2): 377–395.

Kayhan, A. and Titman, S. (2007). Firms' histories and their capital structures. *Journal of Financial Economics*. 83(1): 1-32.

Kent, J.T. and Tyler, D.E. (1996). Constrained M-estimation for multivariate location and scatter. *Annals of Statistics*. 24: 1346–1370.

Kezdi, G. (2004). Robust standard error estimation in fixed-effects panel models. *Hungarian Statistical Review*. 9: 95-116.

Krasker, W.S. and Welsch, R.E. (1982). Efficient bounded-influence regression estimation. *Journal of the American Statistical.* 77: 595–604.

Kunsch, H.R. (1989). The jackknife and the bootstrap for general stationary observations. *Annals of Statistics*. 17: 1217-1241.

Kutner, M.H., Nachtsheim, C.J. and Neter, J. (2004). *Applied Linear Regression Models*. 4th Edition, McGraw Hill, New York, ISBN: 978-0256086010.

Liu, R. (1988). Bootstrap procedure under some non-i.i.d. models. *Annals of Statistics*. 16: 1696–1708.

Liu, R.Y, and Singh, K. (1992). Efficiency and robustness in resampling. *Annals of Statitics*. 20: 370–384.

Liu, R.Y., and Singh, K. (1992). Moving blocks jackknife and bootstrap capture weak dependence. Exploring the Limits of the Bootstrap, ed. by R. LePage and L. Billiard. New York: Wiley.

Lopuhaa, H. (1989). On the relation between S-estimators and M-estimators of multivariate location and covariance. *Annals of Statistics*. 17: 1662–1683.

Lucas, A., Van Dijk, R. and Kloek, T. (2007). Outlier robust GMM estimation of leverage determinants in linear dynamic panel data models. *Unpublished manuscript*, Vrije Universiteit, Amsterdam, the Netherlands.

Lyu, Y. (2015). Detection of outliers in panel data of intervention effects model based on variance of remainder disturbance. *Mathematical Problems in Engineering*. vol. 2015, Article ID 902602, 12 pages, 2015. https://doi.org/10.1155/2015/902602.

MacKinnon, J. and Webb, M. (2015). Wild bootstrap inference for wildly different cluster sizes. No 1314, Working Papers, Queen's University, Department of Economics, https://EconPapers.repec.org/RePEc:qed:wpaper:1314.

Maechler, M, Rousseeuw, P., Croux, C., Todorov, V., Ruckstuhl, A., Salibian-Barrera, M., Verbeke, T., Koller, M., Conceicao and Maria Anna di Palma (2018). robustbase: Basic Robust Statistics R package version 0.93-1. URL http://CRAN.R-project.org/package=robustbase

Mallows, C. L. (1975). *On Some Topics in Robustness*. Unpublished memorandum, Bell Telephone Laboratories, Murray Hill, NJ.

Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *Annals of Statistics.* 21: 255-285.

Maronna, R.A. and Yohai, V.J. (1995). The behavior of the Stahel-Donoho robust multivariate estimator. *Journal of the American Statistical Association*. 90: 329-341.

Maronna, R.A. and Yohai, V.J. (1998). Robust estimation of multivariate location and scatter. *Encyclopedia of Statistical Sciences*, Kotz, S., C. Read and D. Banks (Eds.). Wiley-Interscience, New York, ISBN: 10: 0471118362, pp: 589-596.

Maronna, R. A. and Zamar, R. H. (2002). Robust estimates of location and dispersion for high- dimensional datasets. *Technometrics.* 44: 307–317.

Maronna, R.A., Martin, R.D. and Yohai, V.J. (2006). *Robust Statistics Theory and Methods*. New York:John Wiley and Sons.

Maronna, R. A. (1976). Robust M-Estimators of multivariate location and scatter. *The Annals of Statistics*. 4: 51–67.

Miller, S. and Startz, R. (2018). Feasible generalized least squares using machine learning. *SSRN Library*.

Montgomery, D.C., Peck, E.A. and Viving, G.G. (2001). *Introduction to Linear Regression Analysis*. 3rd Edition. New York: John Wiley and Sons.

Mosteller, F. and Tukey , J.W. (1977) . *Data Analysis and Regression*. Reading MA: Addison-Wesley Publishing Company.

Norazan, M.R., Habshah, M. and Imon, A.H.M.R. (2009). Estimating regression coefficients using weighted bootstrap with probability. *WSEAS Transactions on Mathematics.* 8(7): 362–371.

Peña, D. and Prieto, F.J. (2001). Multivariate outlier detection and robust covariance matrix estimation. *Technometrics*. 43(3): 286-310.

Pena, D. and Yohai, V. (1999). A fast procedure for outlier diagnostics in large regression problems. *Journal of the American Statistical Association.* 94(446): 434-445.

Petersen, M.A. (2009). Estimating standard errors in finance panel data sets: comparing approaches. *The Review of Financial Studies.* 22: 435-480.

Preminger, A. and Franck, R. (2007). Forecasting exchange rates: a robust regression approach. *International Journal of Forecasting*. 23(1): 71 – 84.

R Core Team (2014). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Ramsay, J. (1977). A comparative study of several robust estimates of slope, intercept, and scale in linear regression. *Journal of the American Statistical Association.* 72(359): 608-615.

Sohel, R., Habshah, M. and Imon, A.H.M.R (2008). A robust modification of the Goldfeld-Quandt test for the detection of heteroscedasticity in the presence of outliers. *Journal of Mathematics and Statistics*.

Stock, J. H., Watson, M. W. (2003). Introduction to Econometrics. Boston: Addison-Wesley.

Riazoshams, H., Habshah, M. and Sharipov, O. (2010). The performance of robust two-stage estimator in nonlinear regression with autocorrelated error *Communications in Statistics — Simulation and Computation*. 39(6): 1251-1268.

Rocke, D. (1996). Robustness properties of S-estimators of multivariate location and shape in high dimension. *The Annals of Statistics.* 24(3): 1327-1345.

Rousseeuw, P.J. (1985). Multivariate estimation with high breakdown point. *Mathematical and Statistical Applications*. B: 283-297.

Rousseeuw, P.J. (1984). Least median of squares regression. *Journal of the American Statistical Association.* 79: 871–880.

Rousseeuw, P.J. and Leroy, A.M. (1987). *Robust Regression and Outlier Detection*. New York: Wiley.

Rousseeuw P.J. and Van Driessen, K. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics.* 41:212–223.

Rousseeuw, P. and Van Zomeren, B. (1990).Unmasking multivariate outliers and leverage points. *Journal of American Statistical Associations*. 85: 633-639.

Rousseeuw, P.J, Daniels, B., and Leroy, A. (1984) “Applying Robust Regression to Insurance”, *Insurance: Mathematics and Economics*, 3(1):67 – 72.

Rousseeuw, P.J. (1985). Multivariate estimation with high breakdown point. *Mathematical and Statistical Applications*. B: 283-297.

Rousseeuw, P.J. and Hubert, M. (1997). *Recent Development in PROGRESS. In: L1-Statistical Procedures and Related Topics*, Dodge, Y. (Ed.). Vol. 31, Institute of Mathematical Statistics, Hayward, California, pp: 201-214.

Rousseeuw, P.J. and Leroy A. M. (2003). *Robust Regression and Outlier Detection*. John Wiley, New York, 2003.

Ruppert, D. (1992). Computing S Estimators for Regression and Multivariate Location/Dispersion. *Journal of Computational and Graphical Statistics*. 1:253-270.

Ruppert, D. and Simpson, D. G. (1990). Unmasking multivariate outliers and leverage points: Comment. *Journal of the American Statistical Association*. 85:644-646.

Ryan, T. P. (1997). *Modern Regression Methods*. New York: Wiley.

Salibian-Barrera, M. and Yohai, V.J. (2006) “A Fast Algorithm for S-regression Estimates”, *Journal of Computational and Graphical Statistics*. 15(2): 414-427.

Salibian-Barrera, M. and Zamar, R.H. (2002). Bootstrapping robust estimates of regression. *The Annals of Statistics*. 30, 556–582.

Salibian-Barrera, M.; Van Aelst, S. and Willems, G. (2006). PCA based on multivariate MM-estimators with fast and robust bootstrap. *Journal of the American Statistical Association*. 101:1198–1211.

Serfling, R. and Wang, S. (2014). General foundations for studying masking and swamping robustness of outlier identifiers. *Statistical Methodology.* 20:79–90. (Special Issue in memory of Professor Kesar Singh)

Shao, J. (1990). Bootstrap estimation of the asymptotic variances of statistical functionals. *Annals of the Institute of Statistical Mathematics.* 42:737–752.

Shao, J. (1996). Bootstrap model selection. *Journal of the American Statistical Association.* 91(434): 655-65.

Simpson, J. R. (1995). *New Methods and Comparative Evaluations for Robust and Biased-Robust Regression Estimation*. Unpublished Ph.D. thesis, Arizona State University, The United States of America.

Simpson J. R. and Montgomery D. C. (1998). The development and evaluation of alternative Generalized M-estimation techniques. *Communications in Statistics - Simulation and Computation*. 27(4):999–1018.

Simpson, D. G., Ruppert, D. and Carroll, R. J. (1992). On one-step GM estimates and stability of influences in linear regression. *Journal of the American statistical association*. 87: 439-450.

Singh, K. (1998). Breakdown theory for bootstrap quantiles, *The Annals of Statistics.* 26:1719–1732.

Stahel, W.A. (1981). *Breakdown of Covariance estimators*, Research report 31, Fachgruppe für Statistik, Swiss Federal Institute of Technology (ETH):Zürich, (1981).

Stock, J. and Watson,M. (2008). Heteroskedasticity-robust standard errors for fixed effects panel data regression. *Econometrica*. 76(1):155–174.

Stromberg, A.J., Hossjer, O. and Hawkins, D.M. (2000). The least trimmed differences regression estimator and alternatives. *Journal of the American Statistical Association*. 95: 853-864.

Stromberg A.J. (1997). Robust covariance estimates based on resampling. *Journal of Statistical Planning and Inference.* 57:321–334.

Thompson, S. B. (2011). Simple formulas for standard errors that cluster by both firm and time. *Journal of Financial Economics.* 99:1-10.

Tyler, D. E. (1994). Finite sample breakdown points of projection-based multivariate location and scatter statistics. *The Annals of Statistics*. 22:1024-1044.

Uraibi, H.S., Habshah, M., Talib, B.A. and Yousif, J. (2009). Linear regression model selection based on robust bootstrapping technique. *American Journal of Applied Sciences.* 6: 1191-1198.

Venables, W.N. and Ripley, B.D. (2003). *Modern Applied Statistics with S-Plus.* 4th Edition, Springer-Verlag.

Verardi, V. and Wagner, J. (2010). Robust estimation of linear fixed effects panel data models with an application to the exporter productivity premium. *SSRN eLibrary*.

Wagenvoort, R. and Waldmann, R. (2002). On B-robust instrumental variable estimation of the linear model with panel data. *Journal of Econometrics*. 106:297-324.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct tests for heteroscedasticity. *Econometrica*. 48(4):817-838.

Wilcox, R. R. (2005). *Introduction to Robust Estimation and Hypothesis Testing*. 2nd edition. The United States of America: Elsevier Academic Press.

Willems, G. and Van Aelst, S. (2005). Fast and robust bootstrap for LTS, *Computational Statistics and Data Analysis*. 48:703–715.

Woodruff, D.L. and Rocke, D.M. (1994). Computable robust estimation of multivariate location and shape in high dimension using compound estimators. *Journal of the American Statistical Association.* 89:888–896.

Wu, C. (1986): Jackknife, bootstrap and other resampling methods in regression analysis. *The Annals of Statistics*. 14:1261–1295.

Yohai, V.J. (1987). High breakdown point and high efficiency robust estimates for regression. *The Annals of Statistics*. 15: 642-656.

Zaman, A., Rousseeuw, P.J. and Orhan, M. (2001). Econometric applications of high-breakdown robust regression techniques. *Economics Letters*. 71(1):1–8.